Vol.12, No.3, Fall 2023, pp.101-122 ISSN (Print):2322-1682 ISSN (Online):2383-3076

Journal of Geography and Environmental Hazards

RESEARCH ARTICLE DOI:10.22067/geoeh.2023.79402.1296 Open access

Estimation of PM2.5 as a Harmful Environmental Hazard in Tehran by Fusion of MODIS Aerosol Products through a Machine Learning Approach

Ali Mirzaei[a], Hossein Bagheri[b*], Mehran Sattari[b]

[a] *MA in Reomte Sensing, University of Isfahan, Isfahan, Iran*
[b] *Assistant Professor in Geomatics, University of Isfahan, Isfahan, Iran*

Received:14 November 2022 ***Revised: 12 December 2022*** ***Accepted: 15 January 2023***

Abstract

Air pollution is one of the most harmful natural hazards in Tehran metropolitan. Particles with a diameter of less than 2.5 micrometers (PM2.5) as one of the most harmful pollutants have endangered the health of people living in Tehran. One of the PM2.5 estimation techniques is the use of Aerosol Optical Depth (AOD) products derived from satellite observations. Various AOD products are retrieved with different algorithms that do not have the same accuracy and spatial resolution. Due to the differences in many assumptions and approximations adopted in the AOD retrieval process, the generated AOD products involve uncertainties. This issue reduces the accuracy of PM2.5 concentration estimation. The purpose of this study was to investigate the possibility of fusing AOD products obtained from MODIS sensor observations (retrieved by Deep Blue and Dark Target algorithms) to estimate PM2.5 more accurately. First, the performance of different machine learning algorithms in estimating PM2.5 from AOD data was evaluated. As a result, the XGBoost algorithm with the highest performance was selected as the base model for PM2.5 estimation. Then, the AOD products were fused using a weighted averaging based on the retrieval quality of the primary products. Finally, the fused AOD product along with meteorological data were employed to estimate PM2.5 using XGboost. The results demonstrated that the accuracy of PM2.5 estimation from the fused AOD product is better than when the AOD products are used individually (R^2= 0.77, MAE = 7.00 $\frac{\mu g}{m^3}$, RMSE = 9.59 $\frac{\mu g}{m^3}$). Thus, the retrieval quality of AOD products will lead to more accurate estimation of PM2.5 in the end.

Keywords: Aerosol Optical Depth, Data Fusion, MODIS, PM2.5, XGBoost

---

*. Corresponding author: Hossein Bagheri Email: h.bagheri@cet.ui.ac.ir Tel:+989127567529

How to cite this Article: Mirzaei, A., Bagheri, H., & Sattari, M. (2023). Estimation of PM2.5, as a harmful environmental hazard in Tehran by fusion of MODIS aerosol products through a machine learning approach. *Journal of Geography and Environmental Hazards*, *12*(3), 101-122

DOI: 10.22067/geoeh.2023.79402.1296

Geography and Environmental Hazards
Volume 12, Issue 3 - Number 47, Fall 2023
https://geoeh.um.ac.ir
https://doi.org/10.22067/geoeh.2023.79402.1296

جغرافیا و مخاطرات محیطی، سال دوازدهم، شمارهٔ چهل و هفتم، پاییز ۱۴۰۲، صص ۱۲۲-۱۰۱

مقاله پژوهشی

ارائه یک روش مبتنی بر یادگیری ماشین برای تلفیق محصولات آئروسل سنجندهٔ مادیس جهت بهبود تخمین PM2.5، به‌عنوان یکی از مهم‌ترین مخاطرات زیست‌محیطی در شهر تهران

علی میرزایی -کارشناسی ارشد دانشکدهٔ مهندسی عمران و حمل‌ونقل، دانشگاه اصفهان، اصفهان، ایران
حسین باقری[1] -استادیار گروه مهندسی نقشه‌برداری، دانشکده مهندسی عمران و حمل‌ونقل، دانشگاه اصفهان، ایران
مهران ستاری-استادیار گروه مهندسی نقشه‌برداری، دانشکده مهندسی عمران و حمل‌ونقل، دانشگاه اصفهان، ایران

تاریخ دریافت: ۱۴۰۱/۸/۲۳ تاریخ بازنگری: ۱۴۰۱/۹/۲۱ تاریخ تصویب: ۱۴۰۱/۱۰/۲۵

چکیده

افزایش بحران‌های محیط‌زیست در جهان باعث شده است تا در دهه‌های اخیر اهمیت مطالعه در مورد مسائل زیست‌محیطی افزایش یابد. آلودگی هوا در ردیف یکی از زیان‌بارترین مخاطرات طبیعی قرار گرفته است. ذرات معلق در هوا با قطر کمتر از ۲.۵ میکرومتر (PM2.5) یکی از خطرناک‌ترین و مضرترین نوع ذرات در میان آلاینده‌های مختلف هوا برای سلامتی انسان هستند. یکی از تکنیک‌های تخمین PM2.5 استفاده از محصولات عمق نوری آئروسل (AOD) است. محصولات مختلف AOD با الگوریتم‌های متفاوتی بازیابی می‌شوند که دارای دقت و قدرت تفکیک مکانی یکسانی نمی‌باشند. به دلیل تفاوت در فرضیات و تقریبات زیادی که در مراحل بازیابی AOD ها اتخاذ می‌شود، محصولات AOD تولید شده دارای عدم قطعیت هستند. این موضوع، باعث کاهش دقت تخمین غلظت PM2.5 می‌گردد. هدف این مقاله بررسی امکان تلفیق محصولات AOD حاصل از مشاهدات سنجندهٔ MODIS (بازیابی شده توسط

[1] نویسنده مسئول: ۰۹۱۲۷۵۶۷۵۲۹ Email: h.bagheri@cet.ui.ac.ir

نحوه ارجاع به این مقاله:

میرزایی, علی, باقری, حسین, ستاری, مهران. (۱۴۰۲). ارائه یک روش مبتنی بر یادگیری ماشین برای تلفیق محصولات آئروسل سنجنده‌ی مادیس جهت بهبود تخمین PM2.5، به عنوان یکی از مهمترین مخاطرات زیست محیطی در شهر تهران. جغرافیا و مخاطرات محیطی, ۱۲(۳), صص ۱۲۲-۱۰۱

https://doi.org/10.22067/geoeh.2023.79402.1296

الگوریتم‌های Deep Blue و Dark Target) به‌منظور تخمین دقیق‌تر PM2.5 اســـت. در این مطالعه، نخست با انجام آزمایش بر روی الگوریتم‌های مختلف یادگیری ماشین و بررسی عملکرد آن‌ها در تخمین PM2.5 از روی‌داده‌های AOD، الگوریتم XGBoost به‌عنوان الگوریتم پایه در روش تلفیق پیشـــنهادی انتخاب شـــد. ســـپس محصـــول AOD تلفیقی با اســـتفاده از یک روش وزن‌دهی مبتنی بر کیفیت بازیابی محصـولات اولیه، تولید شـد. محصـول تلفیقی به همراه داده‌های هواشـناسـی و الگوریتم XGBoost برای تخمین PM2.5 مورد اســـتفاده قرار گرفت. نتایج حاصـــل از تلفیق نشـــان داد که دقت تخمین PM2.5 محصـول تلفیقی در هر سه شاخص RMSE، MAE و $\mathbf{R^2}$ نسبت به حالت‌هایی که محصـولات به‌صـورت منفرد اسـتفاده شـدند، بهتر اسـت ($\mathbf{R^2}$ =۰.۷۷، MAE = ۷.۰۰ $\frac{\mu g}{m^3}$، RMSE=۹.۵۹ $\frac{\mu g}{m^3}$). علاوه بر افزایش دقت، روش پیشـنهادی سـاده و از نظر محاسـباتی کم‌هزینه اسـت. علاوه بر این در این تحقیق مشـخص گردید، توجه به **نشانگر کیفی بازیابی محصولات AOD، زمینهٔ دستیابی به یک محصول تلفیقی دقیق‌تر را فراهم خواهد کرد.**

کلیدواژه‌ها: عمق نوری آئروسل، تلفیق داده، سنجندهٔ MODIS، PM2.5، الگوریتم XGBoost.

۱- مقدمه

امروزه آلودگی هوا به یکی از مهم‌ترین مخاطرات زیســت‌محیطی تبدیل شــده اســت (Huang & Kuo, 2018). رشد جمعیت شهرنشین، توسعهٔ صنعت، افزایش ترافیک و عدم برنامه‌ریزی مناسب در زمینهٔ کنترل منابع آلودگی هوا ازجمله عواملی هسـتند که منجر به افزایش انتشـار آلاینده‌ها و تشـدید این مشـکل شـده‌اند. از مهم‌ترین این آلاینده‌ها، ذرات معلق[1] (PM) هستند (Luo et al., 2020). در دهه‌های اخیر علاوه بر منابع طبیعی (مثل آتشفشان‌ها، آتش‌سوزی و طوفان‌های گردوخاک)، افزایش منابع غیرطبیعی مانند اسـتفادهٔ زیاد از سـوخت‌های فسـیلی در بخش‌های صـنعتی و حمل‌ونقل نیز انتشار ذرات معلق را تشدید کرده است. این ذرات به دلیل متفاوت بودن خصوصیاتی همچون اندازه و ترکیبات شیمیایی تشکیل دهنده، تأثیرات یکسانی بر محیط‌زیست و سلامتی انسان ندارند. ذرات با قطر آئرودینامیکی کوچک‌تر از ۱۰ میکرومتر را به‌اختصـار PM10 و ذرات با قطر آئرودینامیکی کوچک‌تر از ۲/۵ میکرومتر را به‌اختصـار PM2.5 می‌نامند. افزایش غلظت ذرات PM2.5 علاوه بر تهدید سلامت انسان باعث بروز هزینه‌های اقتصادی و تأثیر منفی بر حاصـــلخیزی خاک، چرخه غذایی، فتوســـنتز گیاهان و منابع آبی می‌شـــود (Liao et al., 2020; Singer, Zobeck, Poberezsky, & Argaman, 2003). بنابراین اندازه‌گیری و پایش میزان غلظت این ذرات امری ضـــروری است.

1 Particulate Matter

یکی از روش‌های معمول در پایش ذرات PM2.5 اندازه‌گیری میزان غلظت آن‌ها با استفاده از ایستگاه‌های زمینی کنترل کیفیت هوا است؛ اما ایستگاه‌های زمینی هزینه احداث و نگهداری بالایی دارند. همچنین، به دلیل کم بودن تعداد این ایستگاه‌ها، امکان تخمین دقیق و پایش تغییرات فضایی غلظت این ذرات خصوصاً در نواحی با تغییرات بالا نظیر شهرها میسر نیست. به‌عنوان یک راه‌حل، از محصولات AOD[1] حاصل از مشاهدات ماهواره‌ای برای تخمین این آلاینده‌ها استفاده می‌شود. مشاهدات ماهواره‌ای علاوه بر پوشش مکانی خوب و پیوستگی زمانی، بسیار مقرون‌به‌صرفه هستند (Anderson et al., 2010). شاخص AOD با تعیین میزان عمق نوری اتمسفر، به‌عنوان یک شاخص مهم در مطالعهٔ میزان کیفیت هوا و غلظت ذرات PM2.5 مورد استفاده قرار می‌گیرد. AOD شاخصی است که میزان جلوگیری از انتشار نور در یک ستون از جو بر اثر حضور ذرات معلق در هوا را اندازه‌گیری می‌کند. بنابراین این شاخص به‌طور مستقیم با میزان غلظت ذرات معلق در جو ارتباط دارد و به‌عنوان یک شاخص کلیدی در مطالعه PM2.5 استفاده می‌شود ((Wei, Chang, Bai, & Gao, 2020) مطالعات زیادی جهت تخمین میزان غلظت ذرات PM2.5 از روی محصولات AOD صورت پذیرفته است (Bagheri, 2022; Hu et al., 2014; Li, 2020; Ni, Cao, Zhou, Cui, & P Singh, 2018; Qi, Li, Karimian, & Liu, 2019). بااین‌حال استفاده از محصولات AOD نیز با چالش‌ها و نواقصی همراه است.

کیفیت داده‌های ورودی خصوصاً داده‌های AOD تأثیر بسیاری در دقت تخمین غلظت PM2.5 دارد. در سال‌های اخیر الگوریتم‌های متنوع بازیابی AOD (همچون الگوریتم‌های Deep Blue، Dark Target، MAIAC[2]، SARA[3] و ...) توسط تیم‌های تحقیقاتی مختلف گسترش یافته‌اند (Hui Xu et al., 2012). الگوریتم‌های مختلف بازیابیAOD همیشه خروجی یکسانی از نظر دقت و قدرت تفکیک مکانی نمی‌دهند و در مناطق مختلف عملکردهای متفاوتی دارند (Tang, Bo, & Zhu, 2016). به‌عنوان‌مثال، محصول (DT) Dark Target برای مناطقی که بازتاب سطح در محدوده مرئی کم است (مانند مناطق دارای پوشش گیاهی) عملکرد بهتری نشان می‌دهد؛ درحالی‌که محصول Deep Blue (DB) در مناطق با سطوح روشن مانند مناطق شهری نتایج دقیق‌تری به دست می‌دهد (Hsu et al., 2013; Remer, Mattoo, Levy, & Munchak, 2013). بااین‌وجود، عدم قطعیت‌های زیادی در محصولات فوق وجود دارد و این عدم قطعیت‌ها از فرضیات و تقریب‌های زیادی که در مراحل بازیابی AOD ازجمله بخش‌های غربالگری ابر، انتخاب مدل بازیابی و تعیین بازتاب سطح اتخاذ می‌شود، ناشی می‌گردند. بدیهی است که عدم قطعیت در محصولات AOD باعث کاهش کیفیت این محصولات و به‌تبع آن کاهش دقت تخمین PM2.5 می‌شود.

1 Aerosol Optical Depth
2 Multi Angle Implementation of Atmospheric Correction
3 Simplified high resolution MODIS Aerosol Retrieval Algorithm

یکی از روش‌های افزایش کیفیت محصـولات مختلف AOD، تلفیق آن‌ها اسـت. به‌طورکلی مدل‌های تلفیق داده، اطلاعات چندین سـنجنده را جهت بهبود دقت و دستیابی به اطلاعات جامع‌تر، ترکیب می‌کنند (Castanedo, 2013; Hall & Llinas, 1997; Khaleghi, Khamis, Karray, & Razavi, 2013; Meng, Jing, Yan, & Pedrycz, 2020; Mirzaei, Bagheri, & Sattari, 2023). تلفیق این محصـــولات منجر بـه رفع عیوب هر یـک و همچنین افزایش توانایی‌های بالقوه آن‌ها می‌شود

(Bagheri, Schmitt, d'Angelo, & Zhu, 2018; Bagheri, Schmitt, & Zhu, 2018; H Xu et al., 2015). در مطالعات گذشـته نیز تلاش‌هایی برای بهبود عملکرد محصـولات AOD صـورت پذیرفته اسـت. به‌عنوان‌مثال، هان و همکاران[1] (۲۰۱۷) محصـــولات AOD ســـنجنده‌های MODIS2 و CALIOP3 را در یک فرآیند دو مرحله‌ای با هم تلفیق کردند. در مرحله اول شرایطی که هرکدام از محصولات این دو سنجنده با کیفیت بهتر عمل می‌کرد، تعیین شد. در مرحله بعد، تلفیق دو محصول براساس نقاط ضعف و قوت آن‌ها انجام شد و دقت به‌طور قابل‌توجهی از ۰.۶۵ و ۰.۵۲ به ترتیب برای محصـول MODIS و CAILOP به ۰.۸۰ برای محصـول تلفیق شـده براسـاس R^2 افزایش یافت (Han, Ding, Ma, & Gong, 2017). همچنین وانگ و همکاران[4] (۲۰۱۹) محصـول عمق نوری آئروسـل سـنجنده MODIS از سکوهای Terra و Aqua که با الگوریتم DB تولید شده بودند را برای بهبود پوشش مکانی تلفیق کردند. دقت محصـول نهایی برابر با دقت محصـول MODIS بود و تعداد پیکسـل‌های دارای مقدار در محصـول تلفیقی۱۲۳ درصد نسبت به محصولات منفرد افزایش یافت. اخیراً در یک مطالعه دیگر به تلفیق محصولات AOD به دست آمده از یک سـنسـور فعال (CALIPSO) و یک سـنسـور غیر فعال (MODIS) پرداخته شـد. برای تلفیق این محصـولات از روش بیزین ماکزیمم آنتروپی اسـتفاده شـد. نتایج بهبود قابل‌توجه پوشـش مکانی را نشـان داد (پوشـش مکانی، برای محصولات منفرد در بازۀ ۰.۳ درصد تا ۷۵.٤ درصد و برای محصول تلفیقی در بازۀ ٥٦.۷ درصد تا ۱۰۰ درصد قرار داشـت) . این در حالی اسـت که دقت محصـول تلفیق شـده براسـاس شـاخص RMSE برابر با ۵۰.۲۰ و دقت محصـولات اولیه به ترتیب ۰.۱۹ و ۰.۱۸ برای محصـولات CALIOP و MODIS بود (Xia et al., 2021). همچنین در مطالعه‌ای در سـال ۲۰۲۲، با بهره‌گیری از مزایای شـبکه‌های بیزین و رویکرد تلفیق محصـولات AOD حاصـل از سـنجنده های CALIOP و MODIS، سـعی در بهبود قدرت تفکیک زمانی محصـول عمودی AOD شـده اسـت. در این تحقیق با بکارگرفتن محصـول MAIAC، مشـاهدات MODIS و محصـول عمودی AOD حاصـل از CALIOP،

1 Han, Ding, Ma, & Gong, 2017
2 Moderate Resolution Imaging Spectroradiometer
3 Cloud-Aerosol Lidar with Orthogonal Polarization
4 Y. Wang et al., 2019

٥ AOD بدون واحد است.

مقادیر AOD در دو لایه ارتفاعی مجزا با قدرت تفکیک مکانی-زمانی MAIAC (به ترتیب یک کیلومتری و روزانه) ارائه شد (Pashayi & Satari, 2022).

علاوه بر کیفیت محصولات AOD، الگوریتم مورد استفاده برای تخمین رابطه AOD با داده‌های هواشناسی و PM2.5 نیز بر دقت تخمین PM2.5 مؤثر است (باقری، ۲۰۲۲). مطالعات زیادی، با استفاده از روش‌های متنوع، جهت تخمین میزان غلظت ذرات PM2.5 از روی محصولات AOD صورت پذیرفته است (Bagheri, 2022; Hu et al., 2014; Li, 2020; Ni et al., 2018; Qi et al., 2019). دسته‌ای از این روش‌ها مبتنی بر یادگیری ماشین می‌باشند که در مطالعات متعددی قابلیت آن‌ها به اثبات رسیده است (Bagheri, 2022, 2023; Chen, Yang, Du, & Huang, 2021; Gogikar, Tripathy, Rajagopal, Paul, & Tyagi, 2021; Hu et al., 2014; Jung, Chen, & Nakayama, 2021; Kianian, Liu, & Chang, 2021; Li, 2020; Ni et al., 2018; Stafoggia et al., 2019; Yang, Yuan, Li, Shen, & Zhang, 2017; You et al., 2016; Zhao et al., 2020). نی و همکاران (۲۰۱۸) از یک شبکه Fully Connected با یک لایه میانی استفاده کردند و رابطه میان PM2.5، محصول DT سنجندهٔ MODIS و داده‌های هواشناسی را تخمین زدند. این شبکه هفت نورون در لایه ورودی، هفت نورون در لایه میانی و یک نورون در لایه خروجی داشت. نتایج، PM2.5 را با دقت ٦٨ درصد براساس شاخص R^2 ارائه کردند. همچنین نبوی و همکاران (۲۰۱۹) از AOD و پارامترهای هواشناسی شامل شش پارامتر بیشینه، کمینه و میانگین دما، زاویهٔ تابش خورشید، میزان دید و سرعت باد برای برآورد غلظت PM2.5 در تهران استفاده کردند. آن‌ها AOD را به کمک ارتفاع لایه تداخل مرزی استاندارد کردند و از الگوریتم درخت‌تصمیم‌گیری برای برآورد غلظت PM2.5 استفاده کردند. در بهترین حالت، دقت نهایی براساس شاخص R^2 به ۰.٦٨ رسید (نبوی و همکاران[1]، ۲۰۱۹). در یک مطالعهٔ دیگر (۲۰۱۹)، برای پیش‌بینی PM2.5،به مقایسهٔ عملکرد سه الگوریتم RF، XGBoost و Fully Connected براساس داده‌های هواشناسی (شامل بیشینه، کمینه و میانگین دمای هوا، رطوبت نسبی، بارش روزانه، میزان دید، سرعت باد، پایداری سرعت باد، فشار هوا و دمای نقطه شبنم) و داده‌های عمق نوری آئروسل پرداخته شد که الگوریتم XGBoost با RMSE برابر ۱۳.۵۸$\frac{\mu g}{m^3}$ بهترین عملکرد را از خود نشان داد (Zamani Joharestani, Cao, Ni, Bashir, & Talebiesfandarani, 2019). یان و همکاران[2] (۲۰۲۰) با سه روش RF، شبکه عصبی و روش XGBoost به تخمین PM2.5 پرداختند. در این تحقیق، روش XGBoost عملکرد بهتری (R^2 برابر ٦٤ درصد) نسبت به دو روش دیگر داشت. باقری (۲۰۲۲) با روش Gradient Boosting و به کمک محصول MAIAC و داده‌های هواشناسی به تهیهٔ نقشه PM2.5 با وضوح‌بالا در شهر تهران پرداخت. در حین این

1 Nabavi et al.
2 Yan et al.

تحقیق، به تأثیر چندین متغیر بر تخمین PM2.5 پرداخته شد. همچنین نشان دادەشد که الگوریتم یادگیری ماشین Gradient Boosting بهترین عملکرد را در میان سایر الگوریتم‌های یادگیری ماشین دارد (R^2 برابر با ٧٤ درصد).

بررسی مطالعات گذشته نشان می‌دهد که یادگیری ماشین به یک موضوع پرطرفدار تحقیقاتی در حوزه تخمین غلظت ذرات PM2.5 از روی داده‌های آئروسل تبدیل شده است. تنوع الگوریتم‌های استفاده شده در مطالعات قبل، اهمیت انتخاب الگوریتم پایه برای تلفیق را دو چندان می‌کند. این مقاله به بررسی تولنایی و دقت الگوریتم‌های مختلف در تخمین غلظت ذرات PM2.5 از داده‌های آئروسل ماهواره‌ای و اطلاعات هواشناسی نیز می‌پردازد. در نتیجه، یک مدل یادگیری ماشین با بهترین عملکرد، به‌عنوان مدل پایه در تلفیق محصولات مختلف AOD حاصل از تصاویر مادیس انتخاب می‌گردد. سپس، تلفیق با استفاده از مدل پایه و شاخص کیفیت بازیابی محصولات AOD به روش میانگین وزن‌دار انجام خواهد شد. این ایده با کم کردن اثر داده‌های بی‌کیفیت و مشارکت بیشتر داده‌های AOD بازیابی شده با دقت بالا در فرآیند تلفیق، منجر به تخمین دقیق‌تر PM2.5 خواهد گردید. به‌طور مشخص هدف اصلی این مطالعه بهبود دقت تخمین PM2.5 از طریق حذف یا کاهش منابع عدم قطعیت موجود در محصولات AOD با استفاده از تلفیق آن‌ها است.

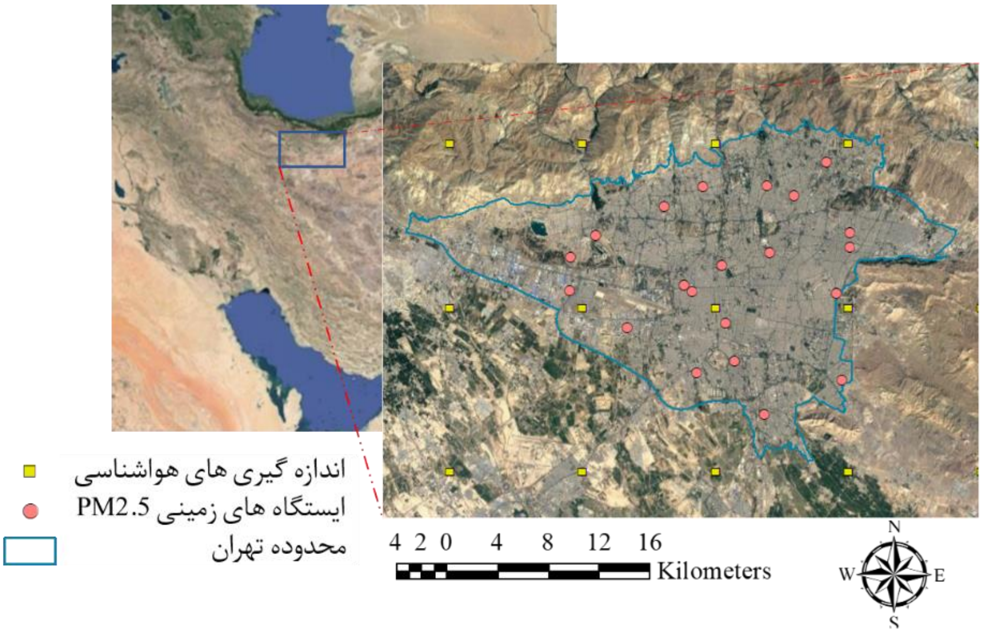
۲- مواد و روش

۲-۱- منطقه موردمطالعه

منطقهٔ موردمطالعه، شهر تهران، پایتخت ایران است که دارای ۱۳.۳ میلیون جمعیت (Habibi, Alesheikh, Mohammadinia, & Sharif, 2017). شکل ۱ موقعیت این منطقه را نشان می‌دهد. این شهر از ۳۵.۳۵ تا ۳۵.٤۸ درجهٔ شمالی و از ۵۱.۱۷ تا ۵۱.۳۳ درجهٔ شرقی گسترش یافته و ۷۵۱ کیلومترمربع مساحت دارد.

۲-۲ داده‌های مورد استفاده

در این مطالعه از محصولات AOD سنجندهٔ MODIS (Deep Blue و Dark Target)، اندازه‌گیری‌های زمینی PM2.5 و داده‌های یک مدل جهانی هواشناسی (European Centre for Medium-Range Weather Forecasts) استفاده گردید. نوع و مشخصات داده‌های مورد استفاده برای تخمین PM2.5 به‌طور خلاصه در جدول ۱ نمایش داده شده‌اند. داده‌ها برای یک بازهٔ زمانی هفت ساله، از ژانویه ۲۰۱۳ تا دسامبر ۲۰۱۹ جمع‌آوری شدند. در ادامه به جزئیات هرکدام از داده‌ها اشاره شده‌است.



شکل ۱- منطقه موردمطالعه، شهر تهران

جدول ۱- داده‌های استفاده شده برای تخمین غلظت PM2.5

داده	توضیح	منبع داده	قدرت تفکیک مکانی	قدرت تفکیک زمانی
AOD	محصولات Deep Blue و Dark Target	سنجنده MODIS	۳ و ۱۰ km	روزانه
Latitude	عرض جغرافیایی ایستگاه‌های اندازه‌گیری PM2.5	AQCC	۱۰ km	روزانه
Longitude	طول جغرافیایی ایستگاه‌های اندازه‌گیری PM2.5	AQCC	۱۰ km	روزانه
Dewpoint temperature (DPT)	دمای نقطه شبنم	ECMWF	۱۰ km	روزانه
Temperature (T)	دما	ECMWF	۱۰ km	روزانه
Boundary layer height (BLH)	ارتفاع لایه مرزی	ECMWF	۱۰ km	روزانه
Surface pressure (SP)	فشار سطحی	ECMWF	۱۰km	روزانه
Leaf area index, high vegetation (Lai hv)	شاخص پوشش برگ، برای مناطق با پوشش متراکم گیاهی	ECMWF	۱۰ km	روزانه
Leaf area index, low vegetation (Lai lv)	شاخص پوشش برگ، برای مناطق با پوشش کم گیاهی	ECMWF	۱۰ km	روزانه
Wind speed (WS)	سرعت باد	ECMWF	۱۰ km	روزانه

داده	توضیح	منبع داده	قدرت تفکیک مکانی	قدرت تفکیک زمانی
Wind direction (WD)	جهت باد	ECMWF	۱۰ km	روزانه
Clear sky direct solar radiation at surface (Cdir)	مقدار تابش رسیده به سطح در شرایط آسمان صاف	ECMWF	۱۰km	روزانه
Downward UV radiation at the surface (Uvb)	مقدار اشعه فرابنفش رسیده به سطح زمین	ECMWF	۱۰km	روزانه
Relative humidity (RH)	رطوبت نسبی	ECMWF	۱۰km	روزانه
Day of year (DOY)	شماره روز سال	-	-	-

۲-۲-۱- محصولات AOD

در این تحقیق از محصولات AOD سنجندهٔ MODIS استفاده شد. این سنجنده AOD را با الگوریتم‌های DT و DB در طول‌موج ۵۵ نانومتر بازیابی می‌کند. سنجندهٔ MODIS بر روی دو سکوی Terra و Aqua قرار دارد. محصولات با قدرت تفکیک زمانی روزانه، به‌صورت رایگان در دسترس هستند (https://earthexplorer.usgs.gov). در کنار تمام محصولات AOD مورد استفاده، یک نشانگر کیفیت بازیابی (Q) نیز ارائه می‌شود. مقادیر Q از ۰ (پایین‌ترین کیفیت) تا ۳ (بالاترین کیفیت) متغیر است. این مقادیر براساس تعداد پیکسل‌های ورودی برای بازیابی AOD، نزدیکی به مناطقی با سطح روشن و مقدار خطای بازیابی، اختصاص داده می‌شوند ("Dark Target aerosol produact user's guid ", 2020).

۲-۲-۲- اندازه‌گیری زمینی PM2.5

همان‌طور که در شکل ۱ نمایش داده شده است، مقدار متوسط روزانه PM2.5 در سطح شهر تهران توسط شرکت کنترل کیفیت هوای تهران[1] (AQCC) در ۲۱ ایستگاه اندازه‌گیری ارائه می‌شود (https://www.iqair.com).

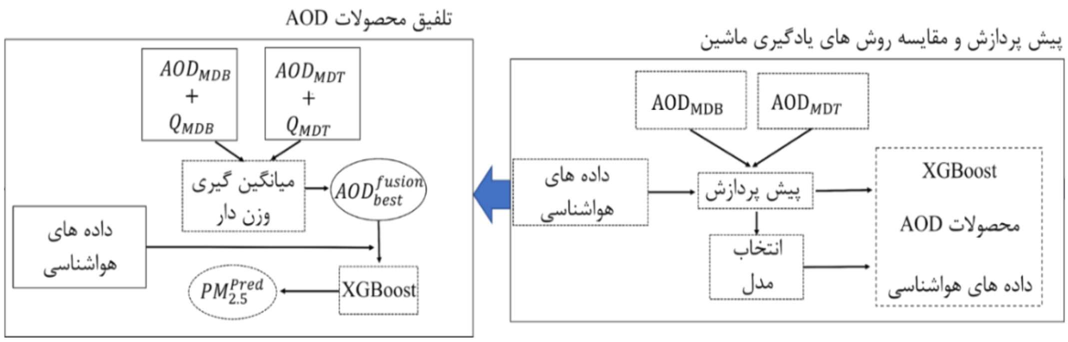
۲-۲-۳- داده‌های هواشناسی

مطالعات مختلفی لزوم استفاده از داده‌های هواشناسی در کنار داده AOD برای تخمین غلظت PM2.5 را روشن ساخته‌اند (Gupta & Christopher, 2009; Ni et al., 2018). برای تهیهٔ داده‌های هواشناسی روزانه از خروجی مدل جهانی ECMWF[2] استفاده شد. این داده‌ها نسبت به داده‌های هواشناسی زمینی در منطقهٔ موردمطالعه، توزیع بهتری دارند. همان‌طور که در جدول ۱ و شکل ۱ مشخص است این داده‌ها با قدرت تفکیک مکانی ۱۰ کیلومتر و قدرت تفکیک زمانی روزانه فراهم شده‌اند.

1 Air Quality Control Company
2 European Centre for Medium-Range Weather Forecasts

۲-۳- روش تحقیق

شکل ۲ روند کلی تحقیق را نشان می‌دهد. نخست، پیش پردازش‌ها بر روی داده‌های خام انجام می‌شود. سپس عملکرد چندین الگوریتم یادگیری ماشین در تخمین PM2.5 با استفاده از محصولات مختلف، بررسی و الگوریتم برتر به‌عنوان الگوریتم پایه در تلفیق محصولات AOD استفاده می‌شود و در نهایت، روش پیشنهادی جهت تلفیق محصولات AOD ارائه می‌شود. از این ساختار کلی می‌توان برای تلفیق دیگر محصولات AOD و نیز در موقعیت‌های مکانی دیگر بهره برد.



شکل ۲- چارچوب پیاده‌سازی شده برای تلفیق محصولات AOD برای تخمین PM2.5

۲-۳-۱- پیش پردازش محصولات AOD

مطالعات گذشته نشان داده‌اند برای تعیین صحت AOD اندازه‌گیری شده می‌توان از مقادیر پیکسل‌های اطراف استفاده کرد (باقری، ۲۰۲۲). به بیان دیگر اختلاف زیاد بین مقادیر اندازه‌گیری شده در یک همسایگی، نشان دهندهٔ حضور ابر، برف و یا دیگر عوارض مخرب است. در این مطالعه نیز به‌منظور استخراج AOD در موقعیت هر ایستگاه زمینی، از یک پنجرهٔ ۳×۳ استفاده می‌شود و درهر پنجره، میانگین مقادیر AOD به‌همراه انحراف معیار آن‌ها محاسبه می‌گردد. چنانچه در هرپنجره انحراف معیار از ۰/۰۲ بیشتر باشد، اندازه‌گیری مذکور غیر قابل اطمینان بوده و حذف می‌گردد. مقدار نهایی AOD در یک پنجره بر اساس میانگین مقادیر قابل اطمینان (پس از بررسی در شرط مذکور) تعیین می‌گردد.

نکته مهم دیگر این است که AOD در یک ستون از جو اندازه‌گیری می‌شود. در حالی که ذرات PM2.5 در نزدیکی سطح حضور دارند. در این خصوص مقادیرAOD قبل از هر پردازشی باید تصحیح شوند. تسای و همکاران[1] (۲۰۱۱) نشان دادند که این تصحیح را می‌توان با تقسیم AOD بر ارتفاع لایه تداخل (Mixing layer

1 Tsai et al.

height) همانند رابطه ۱ بدست آورد(Tsai, Jeng, Chu, Chen, & Chang, 2011; Z. Wang, Chen, Tao, Zhang, & Su, 2010). که در آن nAOD مقدار نرمال شدهٔ AOD است و L_{mix} نیز ارتفاع لایه تداخل را نشان می‌دهد.

$$nAOD = \frac{AOD}{L_{mix}} \qquad (۱)$$

در این مطالعه فرض شد که ذرات آئروسل به‌صورت همگن مخلوط شده و از ارتفاع لایه صرف نظر شد. علاوه براین نبوی و همکاران (۲۰۱۹) نشان دادند که ارتفاع لایه آئروسل و ارتفاع لایه مرزی(Blh) دارای ارتفاع یکسانی بالاتر از لایه‌های مملو از آئروسل هستند در نتیجه، L_{mix} را می‌توان با Blh جایگزین کرد (نبوی و همکاران، ۲۰۱۹). بنابراین، معادله ۱ را می‌توان به‌صورت زیر بازنویسی کرد که آن nAOD مقدار نرمال شدهٔ AOD و Blh نیز ارتفاع لایه مرزی است.

$$nAOD = \frac{AOD}{Blh} \qquad (۲)$$

۲-۳-۲- پیش پردازش PM2.5

در ایستگاه‌های کنترل کیفیت هوای شهر تهران از دستگاه TEOM[1]، برای تعیین غلظت ذرات PM2.5، استفاده می‌شود. این دستگاه قبل از اندازه‌گیری غلظت PM2.5، دمای محیط را تا ۵۰ درجه سانتیگراد بالا می‌برد. این امر باعث می‌شود جرم خشک PM2.5 به‌عنوان غلظت PM2.5 گزارش شود (باقری،۲۰۲۲). مطالعات قبلی نشان داده‌اند رابطه میان AOD و PM2.5 تحت تأثیر رطوبت هوا است (ژانگ و همکاران[2]، ۲۰۱۶). در این تحقیق به کمک فرمول ۳ اثر رطوبت بر PM2.5 اعمال می‌گردد که در آن PM_c، PM تصحیح شده و RH رطوبت نسبی است که بر حسب درصد بیان می‌شود.

$$PM_c = PM(1 - \frac{RH}{100})^{-1} \qquad (۳)$$

۲-۳-۳- پیش پردازش داده‌های هواشناسی

علاوه بر AOD، داده‌های هواشناسی نیز در تخمین PM2.5 نقش مهمی را ایفا می‌کنند. در این مطالعه از داده‌های هواشناسی ECMWF استفاده شد. برای تعیین داده‌های هواشناسی در موقعیت ایستگاه‌های PM2.5، از روش درون‌یابی کریجینگ استفاده شد. هایپرپارامترهای این روش نیز از طریق اعتبار سنجی متقابل[3] تعیین شد (Bagheri, Sadeghian, & Sadjadi, 2014). جدول ۲ دقت درون‌یابی و یکسری اطلاعات آماری برای هر پارامتر هواشناسی را نشان می‌دهد.

1 Tapered Element Oscillating Microbalance
2 Zhang et al.
3 K-Fold Cross Validation

جدول ۲- اطلاعات آماری و دقت درون‌یابی پارامترهای هواشناسی

داده هواشناسی	میانگین	بیشینه مقدار	کمینه مقدار	دقت درون‌یابی (MAE)
Blh (m)	٦٢٢.١	١٩٧١.٢	٢٢.٩	٢٦.٥٢
Cdir ($\frac{J}{m^2}$)	٧٢١٤٢٨.٢	١٠٣٦٤٢٦٨	٣٣٦٦٤٧٨	٦٧٨.٧٤
DPT (K)	٢٧١.٩	٢٨٦.٩	٢٥٢.٥	١.٣٠
T (K)	٢٨٧.٤	٣٠٧.٩	٢٦٢.٤	١.١٧
Lai_hv	١.٨٠	٢.١٩	١.٣٦	٠.٢٢
Lai_lv	٠.٥٠	٠.٥٣	٠.٤٨	٠.٠١
Sp (Pa)	٨٣٣٠٨.٠	٨٦٩٢١.٣	٧٩٧٢٢.٢	٢٤٢٨.٢٤
Uvb ($\frac{J}{m^2}$)	١٢٠٢٦.٤	١٣٨٦٣٧.١	٩٢٣٠٦.٥	١٩٤٦٨١
WD (R)	−٠.٥٢	٢.٠٩	−٢.٤٢	٠.٤٦
WS ($\frac{m}{s}$)	١.٧٤	٤.٧٨	٠.٥٤	٠.٠٧
RH (%)	٧١	٧٩	٦٤	١

۲-۳-٤- مقایسه عملکرد روش‌های یادگیری ماشین وانتخاب مدل پایه برای تلفیق

همانگونه که در فصل مقدمه ذکر گردید، الگوریتم‌های یادگیری ماشینی مختلفی برای تخمین PM2.5 از داده‌های AOD مورد استفاده قرار گرفته‌اند که هرکدام عملکرد متفاوتی را در تحقیقات مختلف بر اساس منطقه موردمطالعه و داده‌های ورودی نشان دادلند؛ بنابراین، در ابتدا لازم است عملکرد چند الگوریتم یادگیری ماشین پرکاربرد، در تخمین PM2.5 با استفاده از محصولات AOD مختلف بررسی گردد. هدف از این کار، یافتن دقیق‌ترین الگوریتم برای استفاده در تلفیق محصولات AOD می‌باشد. در ادامه شرح مختصری از الگوریتم‌های به کاررفته بیان می‌گردد.

یکی از الگوریتم‌های پایه در تخمین PM2.5، مدل رگرسیون خطی چند متغیره می‌باشد که در مطالعات متعددی مورد استفاده قرار گرفته است (Gogikar et al., 2021; Hu et al., 2014; You et al., 2016) . بااین‌حال الگوریتم‌های خطی قادر به حل مسائل غیرخطی پیچیده با ابعاد بالا نیستند. در نتیجه الگوریتم‌های پیشرفته‌تری نظیر الگوریتم رگرسیون بردار پشتیبان[1] (SVR) در حل این مسائل مورد استفاده قرار می‌گیرند (Chen et al., 2021). این الگوریتم مسائل تفکیک ناپذیر در فضایی با ابعاد کم را از طریق یک نگاشت و با بردن به یک فضای جدید با ابعاد بیشتر، با استفاده از یک ابرصفحه تفکیک می‌کند (Chang & Lin, 2011) . علاوه بر الگوریتم ماشین بردار پشتیبان، درخت‌های تصمیم، قابلیت بالایی را در حل مسائل رگرسیون غیرخطی نشان داده‌اند. نسخه پیشرفته‌ای از این الگوریتم‌ها نظیر جنگل تصادفی[2] (RF) در مطالعات بسیاری برای تخمین غلظت PM2.5 از داده‌های AOD مورد استفاده قرار

1 Support Vector Regression
2 Random Forest

گرفته‌اند (Jung et al., 2021; Kianian et al., 2021; Stafoggia et al., 2019). RF با تجمیع[1] کردن درختان تصمیم در برابر بیش برازش مقاوم بوده و ضمن شناسایی داده‌های پرت، می‌تواند اهمیت متغیرهای ورودی را ارزیابی کند(Xiao, Chang, Geng, & Liu, 2018). بااین‌حال، با افزایش درختان و پیچیدگی هر درخت، زمان آموزش و پیش بینی مدل افزایش می‌یابد. الگوریتم XGBoost بر مبنای تقویت[2] درختان تصمیم شکل می‌گیرد. این الگوریتم در برابر بیش از حد یادگیری، مقاوم است و از پردازش موازی در حین آموزش استفاده می‌کند که باعث بالا رفتن سرعت پردازش و دستیابی به دقت بالاتر می‌گردد (Zamani Joharestani et al., 2019). همچنین اخیراً، الگوریتم‌های یادگیری عمیق، عملکرد بسیار موفقی در حوزه‌های گوناگون نظیر رگرسیون و طبقه بندی نشان داده‌اند (Zhu et al., 2020). متداول‌ترین ساختار در حل مسائل رگرسیون با استفاده از داده‌های جدولی، شبکه‌های FC[3] می‌باشند (Bharath Ramsundar & Reza Bosagh Zadeh, 2018).با این وجود‌به‌طورکلی الگوریتم‌های یادگیری عمیق در مواجه با داده‌های جدولی ناهمگون از سایر الگوریتم‌های یادگیری ماشین کلاسیک به ویژه روش‌های مبتنی بر درختان تصمیم، ضعیف‌تر عمل می‌کنند (میرزایی و همکاران[4]، ۲۰۲۳). NODE[5] یک معماری یادگیری عمیق پیشرفته است که برای کار با هر نوع داده جدولی و پوشش این ضعف، طراحی شده است (Popov, Morozov, & Babenko, 2019).

لازم به ذکر است در این تحقیق و در هر آزمایش، ۷۵ درصد داده‌ها به‌عنوان دادۀ آموزشی[6] و ۲۵ درصد داده‌ها به‌عنوان دادۀ آزمایشی[7]، به شکل تصادفی، انتخاب شدند. هایپرپارامترها در تمام الگوریتم‌ها (بجز روش رگرسیون خطی[8] و Fully connected) از طریق روش اعتبار سنجی متقابل، با استفاده از داده‌های آموزشی با انتخاب K =۵ تعیین شدند. همچنین سه شاخص R^2[9]، RMSE[10] و MAE[11] به‌عنوان شاخص دقت، در ارزیابی نتایج، مورد استفاده قرار گرفتند.

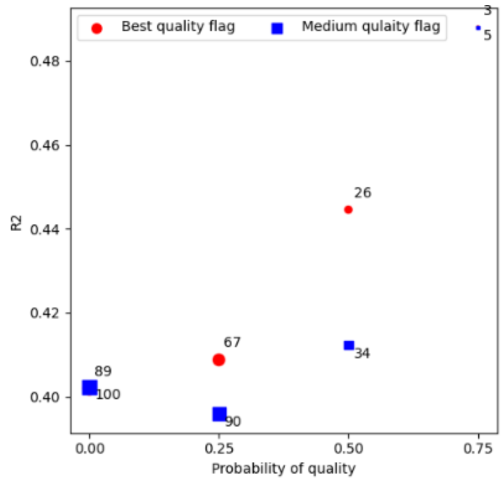
۲-۳-۵- تلفیق محصولات AOD

همان‌طور که قبلاً بیان گردید هدف این پژوهش بهبود تخمین PM2.5 از روی تلفیق داده‌های AOD می‌باشد. بااین‌حال، مطالعات قبلی نیز حاکی از عملکرد متفاوت محصولات AOD در مناطق مختلف‌اند (Liu et al., 2019;

1 Aggregate et al.
2 Boosting
3 Fully connected
4 Mirzaei et al.
5 Neural Oblivious Decision Ensemble
6 Train
7 Test
8 Linear
9 R-squared
10 Root Mean Squared Error
11 Mean Absolute Error

(Sayer et al., 2014; Y. Wang, Yuan, Shen, Zheng, & Zhang, 2020). علت این مسئله را می‌توان در فرضیات استفاده شده برای بازیابی محصولات AOD جستجو کرد (Tang et al., 2016). تلفیق محصولات مختلف AOD منجر به تولید یک محصول خواهد شد که تمام مناطق عملکرد قابل قبولی در تخمین PM2.5 دارد.

در این بخش محصولات AOD به‌صورت مستقیم و بدون پیش‌پردازش با یکدیگر تلفیق می‌شوند. روش پیشنهادی، ساده بوده و بار محاسباتی کمی دارد. چالش اساسی در این نوع تلفیق کاهش نسبت سیگنال به نویز به دلیل استفاده از داده‌های مختلف و متعدد بودن منابع خطا در آنهاست (Bagheri, Schmitt, & Zhu, 2017). مطالعات قبلی نشان داده‌اند که فیلتر محصولات AOD با استفاده از نشانگر کیفیت بازیابی (Q)، باعث بهبود عملکرد آن‌ها در تخمین PM2.5 می‌شوند (Bagheri, 2022; Nabavi, Haimberger, & Abbasi, 2019). شکل ۳ نتایج مطالعهٔ باقری (۲۰۲۲) در خصوص تأثیر اعمال شاخص کیفیت بازیابیAOD محصول MAIAC در تخمین PM2.5 را نشان می‌دهد (Bagheri, 2022). در این مطالعه داده‌های AOD بر اساس دو شرط مربوط به کیفیت بازیابی آن‌ها فیلتر گشتند. حالت یک، فقط شامل شرط ابری نبودن پیکسل مورد نظر و حالت دوم علاوه بر حالت یک، شامل ابری و برفی نبودن پیکسل‌های مجاور است. همان‌طور که در شکل ۳ مشخص است افزایش احتمال حضور پیکسل‌های باکیفیت، باعث افزایش دقت از نظر شاخص R^2 می‌گردد، بااین‌حال فیلتر داده‌ها بر اساس این شاخص، باعث کاهش تعداد نمونه‌ها می‌شود. در این تحقیق نیز برای برای جلوگیری از کاهش نسبت سیگنال به نویز و نیز بهبود دقت محصول نهایی، از شاخص Q برای تلفیق در سطح داده استفاده‌شد. برای این منظور، از میانگین وزن دار AOD ها برای تلفیق آن‌ها استفاده گردید. در ابتدای فرآیند تلفیق، شاخص احتمال حضور مقادیرAOD با کیفیت بالا، در میانگین‌گیری AOD (جزئیات در بخش ۲-۳-۱)، به‌عنوان وزن تعیین می‌گردد. این شاخص، نسبت تعداد پیکسل‌ها با بهترین کیفیت (Q=۳) به تعداد کل پیکسل‌ها در یک پنجره سه در سه اطراف هر ایستگاه را به‌صورت معادله ٤، نشان می‌دهد:



شکل ۳- تأثیر افزایش احتمال حضور AOD های با کیفیت بر اساس شرایط حالت یک و شرایط حالت دو بر تخمین PM2.5 (برای جزئیات بیشتر مراجعه کنید به (Bagheri, 2022))

$$P(AOD_{best}) = \frac{\sum_{i=1}^{9} n}{9} \qquad n = \begin{cases} 1 & if\ Q_i == 3 \\ 0 & else \end{cases} \tag{٤}$$

که در آن Q_i نشان دهندهٔ شاخص کیفیت بازیابی برای پیکسل i ام در یک پنجرهٔ ۳×۳ است و $P(AOD_{best})$ بیانگر احتمال حضور پیکسل با کیفیت، در پنجره مذکور می‌باشد که به‌عنوان وزن پیکسل مربوطه، در تلفیق AOD ها، مورد استفاده قرار می‌گیرد.

پس از تعیین وزن‌ها، AOD تلفیق شده، از میانگین وزن‌دار AOD های ورودی بر اساس احتمال بدست‌آمده در رابطهٔ٤، محاسبه می‌گردد (رابطه ٥):

$$AOD_{best}^{fusion} = \frac{P(AOD_{Terra}^{DT}) \times AOD_{Terra}^{DT} + \cdots + P(AOD_{Aqua}^{DB}) \times AOD_{Aqua}^{DB}}{P(AOD_{Terra}^{DT}) + \cdots + P(AOD_{Aqua}^{DB})} \tag{٥}$$

که در آن اندیس‌های Terra و Aqua، به نام سکوهای هر محصول اشاره دارد. اندیس DT و DB نیز بیانگر الگوریتم بازیابی AOD محصول مورد نظر است. به‌عنوان‌مثال $P(AOD_{Terra}^{DT})$ وزن بدست آمده برای محصول DT سنجندهٔ Terra است یا AOD_{Terra}^{DT} نیز میانگین AODهای بدست آمده در پنجرهٔ ۳×۳ برای محصول DT سنجنده Terraرا نشان می‌دهد. AOD_{best}^{fusion} نیز محصول تفیلق شده با بهترین شرایط کیفی (حاصل از رابطه ٤) می‌باشد. در نهایت AOD تلفیق شده به همراه داده‌های هواشناسی، برای تخمین PM2.5 مورد استفاده قرار می‌گیرد.

۳- نتایج

در این بخش، نتایج مقایسه روش‌های یادگیری ماشین و تلفیق محصولات AOD در تخمین PM2.5 ارائه و مورد بررسی قرار می‌گیرد.

۳-۱- نتایج مقایسه روش‌های یادگیری ماشین

در ابتدا توانایی الگوریتم‌های یادگیری ماشین معرفی شده برای تخمین غلظت PM2.5 با استفاده از محصولات DB و DT، مورد ارزیابی قرار گرفتند. دقت الگوریتم‌ها در تخمین غلظت PM2.5، در جدول ۳ بیان گردیده است. همان‌طور که مشخص است الگوریتم XGBoost در دو شاخص RMSE و R^2 بهترین عملکرد را داشته است. عملکرد الگوریتم XGBoost در تخمین PM2.5 با استفاده از محصول MDT، در شاخص R^2 دو درصد، در شاخص RMSE $\frac{\mu g}{m^3}$ ۰.۱۷ و در شاخص MAE $\frac{\mu g}{m^3}$ ۰.۳۳ بهتر از الگوریتم RF (به‌عنوان بهترین الگوریتم پس از XGBoost) بود. البته XGBoost هنگام استفاده از این محصول، در شاخص MAE و در رقابت با الگوریتم FC اختلاف ناچیزی در حدود $\frac{\mu g}{m^3}$ ۰.۰۳ داشت. از منظر شاخص‌های دقت دیگر و همچنین پارامترهایی مثل زمان آموزش و سهولت در تعیین ابرپارامترها، اولویت با انتخاب الگوریتم XGBoost است. زمانی که از محصول MDB برای تخمین PM2.5 استفاده شد، XGBoost در شاخص R^2 دو درصد و در شاخص RMSE $\frac{\mu g}{m^3}$ ۰.۰۷ نسبت به الگوریتم RF بهتر عمل نمود. درحالی‌که از نظر شاخص MAE، نتایج لندکی ($\frac{\mu g}{m^3}$ ۰.۰۴) بدتر از الگوریتم RF بود. نکتهٔ قابل‌توجه دیگر برتری محسوس XGBoost نسبت به روش‌های یادگیری عمیق آزمایش شده بود. با وجود موفقیت‌های چشم‌گیر این الگوریتم‌ها در زمینه‌های مختلف، این الگوریتم‌ها در مواجه با داده‌های جدولی ناهمگون[1]، عملکرد ضعیف‌تری نسبت به الگوریتم‌های کلاسیک مانند درختان

تصمیم تقویت شده با گرادیان[2]، دارند. همچنین تنظیم ابرپارامترها و زمان آموزش در روش XGBoost بسیار سریع‌تر از روش‌های یادگیری عمیق است (Shwartz-Ziv & Armon, 2022).

جدول ۳- عملکرد روش‌های یادگیری ماشین در تخمین PM2.5

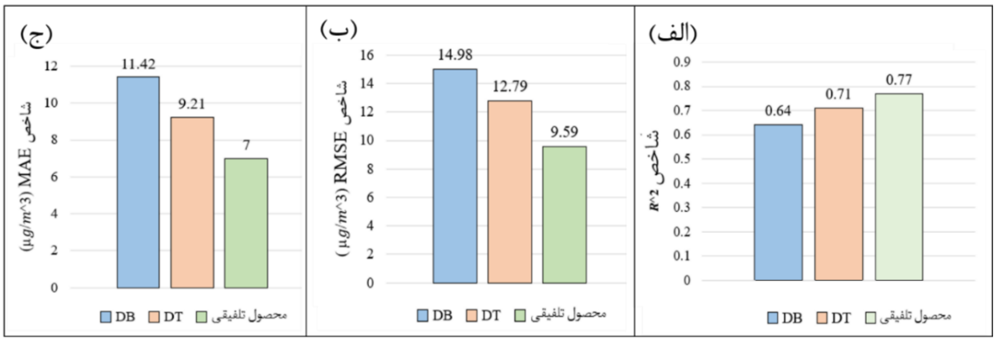
شاخص دقت	R^2		MAE ($\frac{\mu g}{m^3}$)		RMSE ($\frac{\mu g}{m^3}$)	
محصولات	DT	DB	DT	DB	DT	DB
XGBoost	**۰.۷۰**	**۰.۷۶**	۸.۴۱	۸.۴۸	**۱۱.۰۶**	**۱۰.۹۱**
RF	۰.۶۸	۰.۷۴	۸.۷۴	**۸.۴۴**	۱۱.۱۳	۱۱.۰۸
FC	۰.۶۴	۰.۷۰	**۸.۳۸**	۸.۷۰	۱۲.۲۸	۱۰.۹۲

1 heterogeneous tabular data
2 gradient boosted decision trees

شاخص دقت	R^2		MAE ($\frac{\mu g}{m^3}$)		RMSE ($\frac{\mu g}{m^3}$)	
محصولات	DT	DB	DT	DB	DT	DB
NODE	۰.۵۰	۰.۶۵	۱۱.۶۱	۱۰.۲۸	۱۴.۲۹	۱۴.۳۱
SVR	۰.۵۴	۰.۵۸	۱۷.۵۴	۱۴.۶۰	۱۸.۱۷	۲۰.۸۶
Linear	۰.۵۰	۰.۴۹	۱۱.۳۶	۱۱.۸۲	۲۴.۶۸	۱۴.۳۶

۳-۲- نتایج تلفیق

محصولات AOD بر اساس روش توضیح داده شده در بخش ۲-۳-۵ با هم تلفیق گردیدند. شکل ٤ دقت محصول AODتلفیق شده را نسبت به هر یک از محصولات در تخمین PM2.5، نشان می‌دهد. از نظر شاخص R^2، روش پیشنهادی تلفیق و تخمین غلظت PM2.5، شش درصد نسبت به محصول MDT و سیزده درصد نسبت به MDB دقیق‌تر عمل نمود. همچنین نتایج حاصل از تلفیق از نظر شاخص RMSE، $\frac{\mu g}{m^3}$ ۳.۲ نسبت به محصول MDT و $\frac{\mu g}{m^3}$ ۵.۳۹ نسبت به محصول MDB بهبود یافت. عملکرد محصول تلفیقی، از نظر شاخص MAE نیز روش پیشنهادی $\frac{\mu g}{m^3}$ ۲.۲۱ بهتر از MDT و $\frac{\mu g}{m^3}$ ٤.٤۲ بهتر از MDB بود. به‌طورکلی می‌توان چنین نتیجه‌گیری کرد که تلفیق به روش پیشنهادی از نظر دقت، نسبت به هرکدام از محصولات بکار رفته به‌صورت منفرد در تخمین PM2.5 عملکرد بسیار بهتری دارد. اما به ؛ پل شرط استفاده شده برای حذف داده‌های بی‌کیفیت در این روش، برخی اطلاعات که ممکن است در تخمین PM2.5 مفید واقع شوند از دست می‌روند.



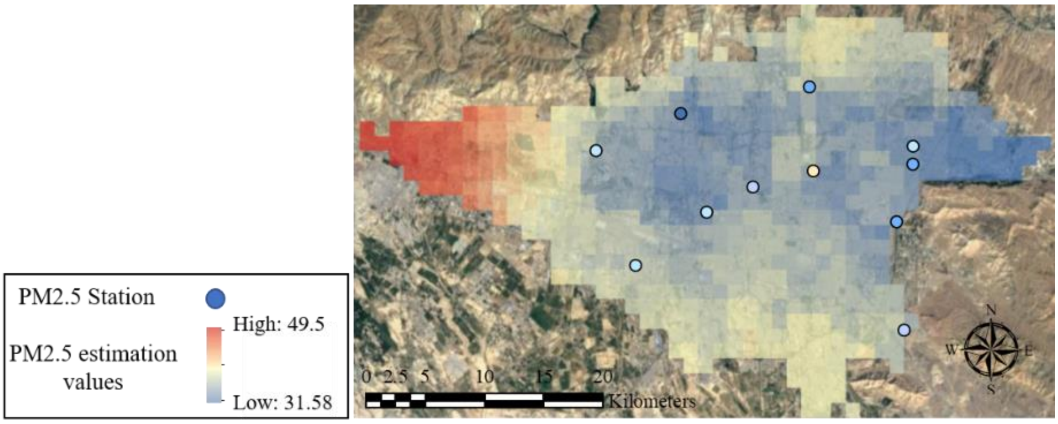
شکل٤- الف) نتایج تلفیق محصولات بر اساس شاخص R^2، ب) نتایج تلفیق محصولات بر اساس شاخص RMSE و ج) نتایج تلفیق محصولات بر اساس شاخص MAE

٤- بحث و جمع‌بندی

در این مطالعه به بررسـی تأثیر تلفیق محصـولات AOD بر بهبود دقت تخمین PM2.5، به‌عنوان یکی از مهم‌ترین مخاطرات زیسـت‌محیطی مخصـوصـاً در مناطق شـهری پرداخته شـد. نخسـت عملکرد الگوریتم‌های مختلف یادگیری ماشین در تخمین PM2.5 ارزیابی شـد. نتایج، نشان دهندهٔ برتری الگوریتم XGBoost بر روی داده‌های مورد استفاده در مقایســـه با روش‌های یادگیری عمیق، بود. ســـپس الگوریتم XGBoost، به‌عنوان الگوریتم پایه در روش تلفیق پیشـنهادی اسـتفاده گردید. روش پیشـنهادی بر اسـاس میانگین وزن‌دار مبتنی بر وزن‌های حاصـل از شـاخص کیفیت بازیابی ارائه شـده در محصـولات AOD، صـورت گرفت. روش مذکور با کم کردن تأثیر مشـارکت AOD های کم کیفیت و نیز افزایش تأثیر AOD های باکیفیت، باعث بهبود قابل‌توجه از نظر دقت نسـبت به محصـولات اولیه برای تخمین PM2.5، شد که در قسمت نتایج ارائه گردید.

همچنین جهت نشان دادن عملکرد مدل پیشنهادی، نقشه PM2.5 حاصل تلفیق داده‌های مختلف AOD در سطح منطقه موردمطالعه برای یک روز ترسیم گردید. برای ترسیم این نقشه، از اندازه‌گیری‌های زمینی PM2.5 و تخمین‌های بدسـت آمده از روش پیشـنهادی و روش درون‌یابی وزن‌دهی معکوس[1] بر اسـاس فاصـله، اسـتفاده شـد. شـکل ۵ نقشـهٔ PM2.5 روز چهارم تیر ماه سال ۱۳۹۵ را نشان می‌دهد. بر اساس PM2.5 اندازه‌گیری شده توسط ایستگاه‌های زمینی در این روز، شاخص کیفیت هوا[2] در دسته متوسط[3] قرار گرفته است. در این روز تنها ۱۱ ایستگاه از ۲۱ ایستگاه زمینی به دلیل مشـکلات فنی، غلظت PM2.5 را ثبت کرده بودند. بر اسـاس شـکل ۵، غلظت PM2.5 با اسـتفاده از روش پیشـنهادی در مناطق مرکزی مناطق تحت پوشش ایستگاه‌های زمینی، بین$\frac{\mu g}{m^3}$ ۳۱ تا $\frac{\mu g}{m^3}$ ٤۹ تخمین زده‌شده است که با AQI اعلام‌شـده از سـوی شـرکت کنترل کیفیت هوای تهران (بین $\frac{\mu g}{m^3}$ ١٦ تا $\frac{\mu g}{m^3}$ ٤١) مطابقت دارد. علت این اختلاف اندک را نیز می‌توان در توزیع نامناسـب ایسـتگاه‌های زمینی اندازه‌گیری PM2.5 جسـتجو کرد. همان‌طور که در شـکل مشـخص اسـت، در منطقه غرب تهران، به دلیل حضـور کارخانه‌ها و منابع آلودگی و نیز وزش باد از سـمت غرب به شـرق، غلظت ذرات PM2.5 بالاسـت. این در حالی اسـت که هیچ ایسـتگاه اندازه‌گیری زمینی در این منطقه وجود ندارد.

1 Inverse Distance Weighting
2 Air Quality Control
3 Moderate



شکل۵- نقشهٔ روزانه PM2.5 تهیه‌شده به کمک تلفیق محصولات مختلف AOD.

تقدیر و تشکر

نویسندگان از سازمان‌هایی که داده‌های خام را به‌صورت رایگان به اشتراک می‌گذارند ازجمله ناسا، برای داده‌های AOD سنجنده MODIS، شرکت کنترل کیفیت هوای تهران برای داده‌های زمینی PM2.5 و مرکز پیش‌بینی هواشناسی متوسط مقیاس اروپا برای داده‌های هواشناسی نهایت تشکر و قدردانی را دارند.

کتابنامه

Anderson, L. O., Malhi, Y., Aragão, L. E., Ladle, R., Arai, E., Barbier, N., & Phillips, O. (2010). Remote sensing detection of droughts in Amazonian forest canopies. *New Phytologist, 187*(3), 733-750. doi:https://doi.org/10.1111/j.1469-8137.2010.03355.x

Bagheri, H. (2022). A machine learning-based framework for high resolution mapping of PM2.5 in Tehran, Iran, using MAIAC AOD data. *Advances in Space Research, 69*(9), 3333-3349. doi:https://doi.org/10.1016/j.asr.2022.02.032

Bagheri, H. (2023). Using deep ensemble forest for high-resolution mapping of PM2. 5 from MODIS MAIAC AOD in Tehran, Iran. *Environmental Monitoring and Assessment*, *195*(3), 377. doi:https://doi.org/10.1007/s10661-023-10951-1

Bagheri, H., Sadeghian, S., & Sadjadi, S. Y. (2014). The Assessment of using an Intelligent Algorithm for the Interpolation of Elevation in the DTM Generation. *Photogrammetrie-Fernerkundung-Geoinformation*, 197-208. doi:https://doi.org/0.1127/1432-8364/2014/0220

Bagheri, H., Schmitt, M., d'Angelo, P., & Zhu, X. X. (2018). A framework for SAR-optical stereogrammetry over urban areas. *ISPRS journal of photogrammetry and remote sensing, 146*, 389-408. doi:https://doi.org/10.1016/j.isprsjprs.2018.10.003

Bagheri, H., Schmitt, M., & Zhu, X. X. (2017). Uncertainty assessment and weight map generation for efficient fusion of TanDEM-X and Cartosat-1 DEMs. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *42*(1/W1)), 433-439. doi:https://doi.org/10.5194/isprs-archives-XLII-1-W1-433-2017

Bagheri, H., Schmitt, M., & Zhu, X. X. (2018). Fusion of TanDEM-X and Cartosat-1 elevation data supported by neural network-predicted weight maps. *ISPRS journal of photogrammetry and remote sensing*, *144*, 285-297. doi:https://doi.org/10.1016/j.isprsjprs.2018.07.007

Bharath Ramsundar & Reza Bosagh Zadeh. (2018). *TensorFlow for Deep Learning* (Rachel Roumeliotis & Alicia Young Ed. first ed. Vol. 16802 KB).

Castanedo, F. (2013). A review of data fusion techniques. *The Scientific World Journal, 2013*, 704504. doi:https://doi.org/10.1155/2013/704504

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST), 2*(3), 1-27. doi:https://dl.acm.org/ doi/10.1145/ 1961189.1961199

Chen, N., Yang, M., Du, W., & Huang, M. (2021). PM2.5 estimation and spatial-temporal pattern analysis based on the modified support vector regression model and the 1 km resolution MAIAC AOD in Hubei, China. *ISPRS International Journal of Geo-Information, 10*(1), 31. doi:https://doi.org/10.3390/ijgi10010031

Dark Target aerosol produact user's guid (2020). In: NASA.

Gogikar, P., Tripathy, M. R., Rajagopal, M., Paul, K. K., & Tyagi, B. (2021). PM2.5 estimation using multiple linear regression approach over industrial and non-industrial stations of India. *Journal of Ambient Intelligence and Humanized Computing, 12*(2), 2975-2991. doi:https://doi.org/10.1007/s12652-020-02457-2

Gupta, P., & Christopher, S. A. (2009). Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: Multiple regression approach. *Journal of Geophysical Research: Atmospheres, 114*(D14). doi:https://doi.org/10.1029/2008JD011496

Habibi, R., Alesheikh, A. A., Mohammadinia, A., & Sharif, M. (2017). An assessment of spatial pattern characterization of air pollution: A case study of CO and PM2.5 in Tehran, Iran. *ISPRS International Journal of Geo-Information, 6*(9), 270. doi: https://doi.org/10.3390/ijgi6090270

Hall, D. L., & Llinas, J. (1997). An introduction to multisensor data fusion. *Proceedings of the IEEE, 85*(1), 6-23. doi:https://doi.org/10.1109/5.554205

Han, B., Ding, H., Ma, Y., & Gong, W. (2017). Improving retrieval accuracy for areosol optical depth by fusiom of MODIS and CALIOP data. *Tehnicki vjesnik/Technical Gazette, 24*(3). doi:https://doi.org/10.17559/TV-20160429044233

Hsu, N., Jeong, M. J., Bettenhausen, C., Sayer, A., Hansell, R., Seftor, C., . . . Tsay, S. C. (2013). Enhanced Deep Blue aerosol retrieval algorithm: The second generation. *Journal of Geophysical Research: Atmospheres, 118*(16), 9296. doi:https://doi.org/10.1002/jgrd.50712

Hu, X., Waller, L. A., Lyapustin, A., Wang, Y., Al-Hamdan, M. Z., Crosson, W. L., . . . Puttaswamy, S. J. (2014). Estimating ground-level PM2.5 concentrations in the Southeastern United States using MAIAC AOD retrievals and a two-stage model. *Remote Sensing of Environment, 140*, 220-232. doi:https://doi.org/10.1016/j.rse.2013.08.032

Huang, C.-J., & Kuo, P.-H. (2018). A deep CNN-LSTM model for particulate matter (PM2. 5) forecasting in smart cities. *Sensors, 18*(7), 2220. doi:https://doi.org/10.3390/s18072220

Jung, C.-R., Chen, W.-T., & Nakayama, S. F. (2021). A national-scale 1-km resolution PM2.5 estimation model over japan using maiac aod and a two-stage random forest model. *Remote Sensing, 13*(18), 3657. doi:https://doi.org/10.3390/rs13183657

Khaleghi, B., Khamis, A., Karray, F. O., & Razavi, S. N. (2013). Multisensor data fusion: A review of the state-of-the-art. *Information Fusion, 14*(1), 28-44. doi:https://doi.org/ 10.1016/ j.inffus. 2011.08.001

Kianian, B., Liu, Y., & Chang, H. H. (2021). Imputing satellite-derived aerosol optical depth using a multi-resolution spatial model and random forest for PM2.5 prediction. *Remote Sensing, 13*(1), 126. doi:https://doi.org/10.3390/rs13010126

Li, L. (2020). A robust deep learning approach for spatiotemporal estimation of satellite AOD and PM2.5. *Remote Sensing, 12*(2), 264. doi:https://doi.org/10.3390/rs12020264

Liao, Q., Jin, W., Tao, Y., Qu, J., Li, Y., & Niu, Y. (2020). Health and Economic Loss Assessment of PM2. 5 Pollution during 2015–2017 in Gansu Province, China. *International Journal of Environmental Research and Public Health, 17*(9), 3253. doi:https:// doi.org/ 10.3390/ ijerph1 7093253

Liu, N., Zou, B., Feng, H., Wang, W., Tang, Y., & Liang, Y. (2019). Evaluation and comparison of multiangle implementation of the atmospheric correction algorithm, Dark Target, and Deep Blue aerosol products over China. *Atmospheric Chemistry and Physics, 19*(12), 8243-8268. doi:https://doi.org/10.5194/acp-19-8243-2019

Luo, H., Guan, Q., Lin, J., Wang, Q., Yang, L., Tan, Z., & Wang, N. (2020). Air pollution characteristics and human health risks in key cities of northwest China. *Journal of Environmental Management, 269*, 110791. doi:https:// doi.org/ 10.1016/ j.jenvman. 2020. 110791

Meng, T., Jing, X., Yan, Z., & Pedrycz, W. (2020). A survey on machine learning for data fusion. *Information Fusion, 57*, 115-129. doi:https://doi.org/10.1016/j.inffus.2019.12.001

Mirzaei, A., Bagheri, H., & Sattari, M. (2023). Data level and decision level fusion of satellite multi-sensor AOD retrievals for improving PM2. 5 estimations, a study on Tehran. *Earth Science Informatics*, *16*(1), 753-771. doi:http://dx.doi.org/10.48550/arXiv.2302.10278

Nabavi, S. O., Haimberger, L., & Abbasi, E. (2019). Assessing PM2.5 concentrations in Tehran, Iran, from space using MAIAC, deep blue, and dark target AOD and machine learning algorithms. *Atmospheric Pollution Research, 10*(3), 889-903. doi:https:// doi.org/ 10.1016/ j.apr.2018.12.017

Ni, X., Cao, C., Zhou, Y., Cui, X., & P Singh, R. (2018). Spatio-temporal pattern estimation of PM2. 5 in Beijing-Tianjin-Hebei Region based on MODIS AOD and meteorological data using the back propagation neural network. *Atmosphere, 9*(3), 105. doi:https:// doi.org/ 10.3390/atmos9030105

Pashayi, M., & Satari, M. (2022). Improvement of spatial-temporal resolution of aerosol profile by using multi-source satellite data over the Persian Gulf. *Atmospheric Environment*, 119410. doi:https://doi.org/10.1016/j.atmosenv.2022.119410

Popov, S., Morozov, S., & Babenko, A. (2019). Neural oblivious decision ensembles for deep learning on tabular data. *arXiv preprint arXiv:1909.06312*. doi:https://doi.org/ 10.48550/ arXiv.1909.06312

Qi, Y., Li, Q., Karimian, H., & Liu, D. (2019). A hybrid model for spatiotemporal forecasting of PM2.5 based on graph convolutional neural network and long short-term memory. *Science of the Total Environment, 664*, 1-10. doi:https://doi.org/10.1016/j.scitotenv.2019.01.333

Remer, L., Mattoo, S., Levy, R., & Munchak, L. (2013). MODIS 3 km aerosol product: algorithm and global perspective. *Atmospheric Measurement Techniques, 6*(7), 1829-1844. doi:https://doi.org/10.5194/amt-6-1829-2013

Sayer, A., Munchak, L., Hsu, N., Levy, R., Bettenhausen, C., & Jeong, M. J. (2014). MODIS Collection 6 aerosol products: Comparison between Aqua's Deep Blue, Dark Target, and “merged” data sets, and usage recommendations. *Journal of Geophysical Research: Atmospheres, 119*(24), 13,965-913,989. doi:https://doi.org/10.1002/2014JD022453

Shwartz-Ziv, R., & Armon, A. (2022). Tabular data: Deep learning is not all you need. *Information Fusion, 81*, 84-90. doi:https://doi.org/10.1016/j.inffus.2021.11.011

Singer, A., Zobeck, T., Poberezsky, L., & Argaman, E. (2003). The PM10 and PM2.5 dust generation potential of soils/sediments in the Southern Aral Sea Basin, Uzbekistan. *Journal of Arid Environments, 54*(4), 705-728. doi:https://doi.org/10.1006/jare.2002.1084

Stafoggia, M., Bellander, T., Bucci, S., Davoli, M., De Hoogh, K., De'Donato, F., . . . Renzi, M. (2019). Estimation of daily PM10 and PM2.5 concentrations in Italy, 2013–2015, using a

spatiotemporal land-use random-forest model. *Environment International, 124,* 170-179. doi:https://doi.org/10.1016/j.envint.2019.01.016

Tang, Q., Bo, Y., & Zhu, Y. (2016). Spatiotemporal fusion of multiple-satellite aerosol optical depth (AOD) products using Bayesian maximum entropy method. *Journal of Geophysical Research: Atmospheres, 121*(8), 4034-4048. doi:https://doi.org/10.1002/2015JD024571

Tsai, T.-C., Jeng, Y.-J., Chu, D. A., Chen, J.-P., & Chang, S.-C. (2011). Analysis of the relationship between MODIS aerosol optical depth and particulate matter from 2006 to 2008. *Atmospheric Environment, 45*(27), 4777-4788. doi:https:// doi.org/ 10.1016/ j.atmosenv. 2009.10. 006

Wang, Y., Yuan, Q., Shen, H., Zheng, L., & Zhang, L. (2020). Investigating multiple aerosol optical depth products from MODIS and VIIRS over Asia: Evaluation, comparison, and merging. *Atmospheric Environment, 230*, 117548. doi:https:// doi.org/ 10.1016/ j.atmosenv. 2020. 117548

Wang, Z., Chen, L., Tao, J., Zhang, Y., & Su, L. (2010). Satellite-based estimation of regional particulate matter (PM) in Beijing using vertical-and-RH correcting method. *Remote Sensing of Environment, 114*(1), 50-63. doi:https://doi.org/10.1016/j.rse.2009.08.009

Wei, X., Chang, N.-B., Bai, K., & Gao, W. (2020). Satellite remote sensing of aerosol optical depth: Advances, challenges, and perspectives. *Critical Reviews in Environmental Science and Technology, 50*(16), 1640-1725. doi:https://doi.org/10.1080/10643389.2019.1665944

Xia, X., Zhao, B., Zhang, T., Wang, L., Gu, Y., Liou, K.-N., . . . Huang, Y. (2021). Satellite-Derived Aerosol Optical Depth Fusion Combining Active and Passive Remote Sensing Based on Bayesian Maximum Entropy. *IEEE Transactions on Geoscience and Remote Sensing, 60*, 1-13. doi:https://doi.org/10.1109/TGRS.2021.3051799

Xiao, Q., Chang, H. H., Geng, G., & Liu, Y. (2018). An ensemble machine-learning model to predict historical PM2.5 concentrations in China from satellite data. *Environmental science & technology, 52*(22), 13260-13269. doi:https://doi.org/10.1021/acs.est.8b02917

Xu, H., Guang, J., Xue, Y., De Leeuw, G., Che, Y., Guo, J., . . . Wang, T. (2015). A consistent aerosol optical depth (AOD) dataset over mainland China by integration of several AOD products. *Atmospheric Environment, 114*, 48-56. doi:https://doi.org/ 10.1016/ j.atmosenv. 2015. 05.023

Xu, H., Xue, Y., Guang, J., Li, Y., Yang, L., Hou, T., . . . Chen, Z. (2012). *A semi-empirical optical data fusion technique for merging aerosol optical depth over China.* Paper presented at the 2012 IEEE International Geoscience and Remote Sensing Symposium.

Yang, Q., Yuan, Q., Li, T., Shen, H., & Zhang, L. (2017). The relationships between PM2.5 and meteorological factors in China: seasonal and regional variations. *International Journal of Environmental Research and Public Health, 14*(12), 1510. doi:https://doi.org/ 10.3390%2 Fijerph14121510

You, W., Zang, Z., Zhang, L., Li, Y., Pan, X., & Wang, W. (2016). National-scale estimates of ground-level PM2.5 concentration in China using geographically weighted regression based on 3 km resolution MODIS AOD. *Remote Sensing, 8*(3), 184. doi:https:// doi.org/ 10.3390/rs8030184

Zamani Joharestani, M., Cao, C., Ni, X., Bashir, B., & Talebiesfandarani, S. (2019). PM2.5 prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data. *Atmosphere, 10*(7), 373. doi:https://doi.org/10.3390/atmos10070373

Zhao, C., Wang, Q., Ban, J., Liu, Z., Zhang, Y., Ma, R., . . . Li, T. (2020). Estimating the daily PM2.5 concentration in the Beijing-Tianjin-Hebei region using a random forest model with a 0.01°× 0.01° spatial resolution. *Environment International, 134*, 105297. doi:https:// doi.org/ 10.1016/j.envint.2019.105297

Zhu, X. X., Hu, J., Qiu, C., Shi, Y., Kang, J., Mou, L., . . . Huang, R. (2020). So2Sat LCZ42: A benchmark data set for the classification of global local climate zones [Software and Data Sets]. *IEEE Geoscience and Remote Sensing Magazine, 8*(3), 76-89. doi:https://d oi.org/ 10.1109/ MGRS.2020.2964708